A (Proposed) User Study on MT-Enabled Scanning

Marianna J. Martindale, iSchool

Advisor: Marine Carpuat, Dept. of Computer Science

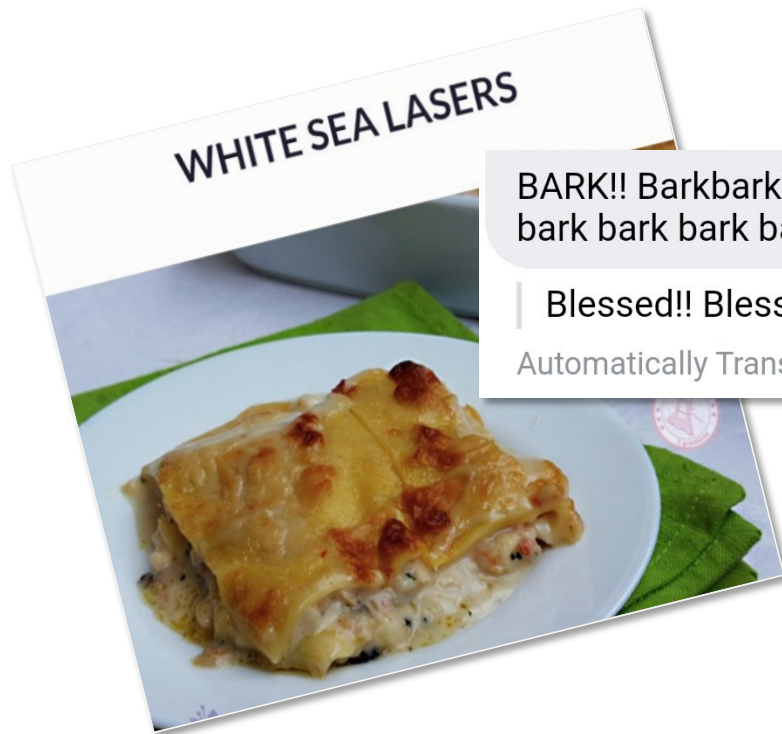University of Maryland, College Park

COLLEGE OF
INFORMATION
STUDIES

UMIACS
University of Maryland Institute for Advanced Computer Studies

MT Errors Happen

(But thankfully not THAT often...)

MT Error Impact

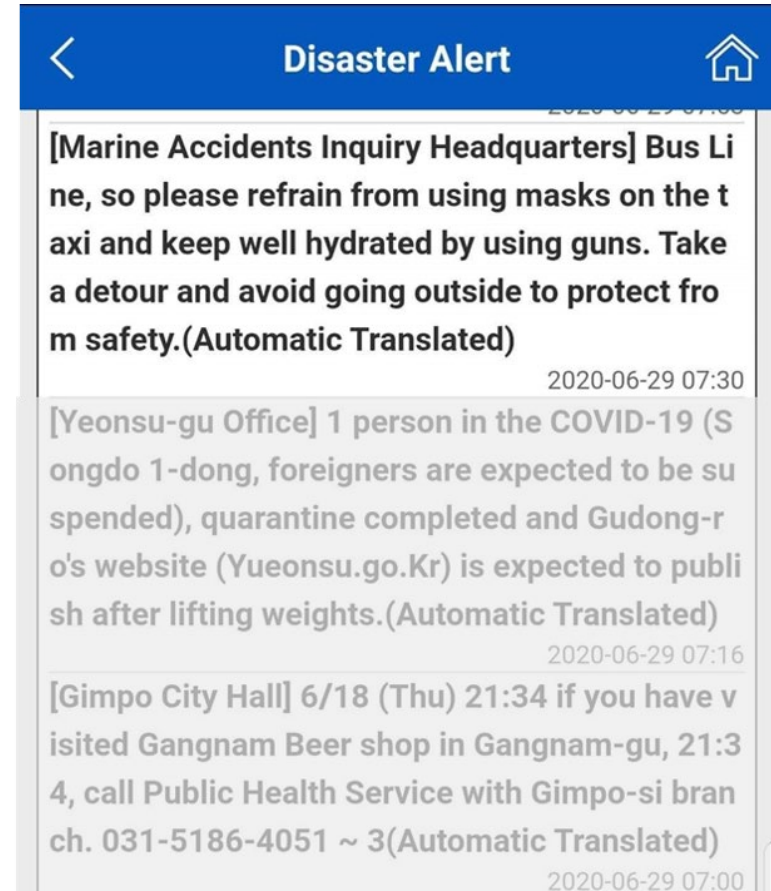**Depends on:**

◦ Severity

  ◦ How wrong is it?

◦ Believability (in context)

  ◦ Laughing? Confused? Convinced?

◦ **Actual use case**

  ◦ Will users take action? What kind?

Disaster Alert

[Marine Accidents Inquiry Headquarters] Bus Line, so please refrain from using masks on the taxi and keep well hydrated by using guns. Take a detour and avoid going outside to protect from safety.(Automatic Translated)
2020-06-29 07:30

[Yeonsu-gu Office] 1 person in the COVID-19 (Songdo 1-dong, foreigners are expected to be suspended), quarantine completed and Gudong-ro's website (Yueonsu.go.Kr) is expected to publish after lifting weights.(Automatic Translated)
2020-06-29 07:16

[Gimpo City Hall] 6/18 (Thu) 21:34 if you have visited Gangnam Beer shop in Gangnam-gu, 21:34, call Public Health Service with Gimpo-si branch. 031-5186-4051 ~ 3(Automatic Translated)
2020-06-29 07:00

Intelligence Analysis MT Use Cases

Like many assimilation use cases
- High volume of foreign language text
- Impractical to translate everything
- Monolingual domain experts use MT to triage

Unique risks and regulations

Personally relevant: 20 years USG MT experience

Use Case: Intelligence Analysis

High level workflow

Intelligence analysts receive trove of foreign language documents

1. Scanning: Identify relevant documents

   Assimilation use case

2. Produce official translation of relevant documents

   Dissemination use case

3. Reporting: Analyze and write report(s)

   (Problematic) Assimilation use case

Scanning Use Case

MT-enabled analysts use MT to get the gist of foreign language documents

- Identify relevant documents and "NTR" (nothing to report) documents
  - Relevance judgment task

- Pass relevant documents to language analysts to translate
  - Often with a contextual note (e.g., "I believe this is a progress report on the HIGH NOON project")
    - Comprehension task

- **Currently acceptable use case and focus of my user study**

Scanning Use Case

Types of error
- **False negative**: relevant document discarded
  - Omits or mistranslates critical information
  - Correct keywords not believable/recognizable in context

- **False positive**: irrelevant document sent for translation
  - Mistranslation or hallucination produces keywords believable in context

"I want to go to lunch at noon, but I generally have to"

"I want to go to lunch, but I have to brief HIGH NOON to the General"

"Want lunch but general high/very noon talk with/about"

Reporting Use Case

Reporting directly off MT output requires deep comprehension

◦ Not simply a binary task = less room for error
  ◦ Example: Errors in numbers or units
    ◦ No effect on scanning
    ◦ Big effect on reporting accuracy!

◦ **Not currently acceptable but tempting**
  ◦ Process more foreign language material
  ◦ Neural MT often looks good enough to use

Not focus of user study, but results may have implications

Former Egyptian President Hosni Mubarak died at the age of five

Former Egyptian President Hosni Mubarak passed away on Tuesday, March 23, at the age of eight. News of Mr Mubarak's death was reported on Egyptian state

Top

User Study Research Questions

Goal: Establish a baseline

- **How accurately can analysts scan short documents w/ MT?**
  - Relevance judgment
  - Comprehension
- **How confident are they in those judgments?**
- **Is their confidence justified by their accuracy?**

Goal: Evaluate Interventions

- **Does intervention reduce how often analyst is misled?**
- **Does it help the analyst calibrate their confidence?**

Interventions

Goals

◦ Raise **prominence** of potential errors

◦ Help users **interpret** them more accurately

◦ Practical in USG environment today
  ◦ Off-the-shelf technology

Intervention A: Two MT Outputs

Intuition:
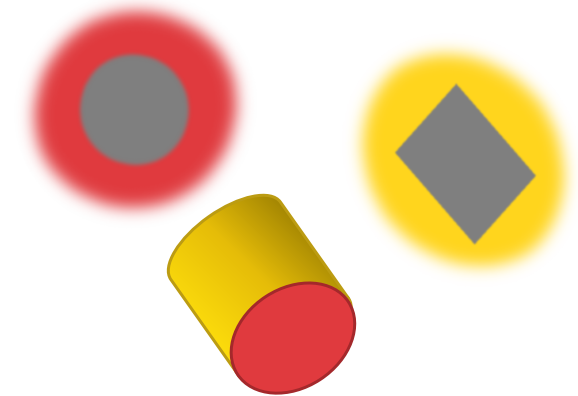◦ Users see and compare outputs
◦ Meaning differences ?= possible mistakes
◦ Common meaning despite disfluency ?= accurate
◦ Anecdotally, users like it!

Related Work:
◦ Shown to be effective in communication use cases
  (Xu et al., 2014; Gao et al., 2015)

Caveat:
◦ Which will they pick when outputs disagree?

Intervention B: Rule-Based MT

Motrans rule-based MT (RBMT)

Intution:

◦ Scan for keywords and compare with NMT

(like intervention A)

◦ Less fluent/comprehensible but more interpretable

◦ Freely available to USG through CyberTrans

Caveat:

◦ Users may ignore it due to lack of fluency/comprehensibility

It's not so ridiculous, I'm looking forward to it more.

MT1

It's not worth a lot of fun analysis. I'm going to make it more than that.

MT2

Analysis very ridiculous without value I don't want to criticize it more from this

RBMT*

*Not actual Motrans output

User Study Scenario
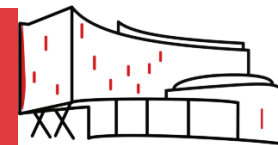
Language: Persian/Farsi

Given chat conversations
◦ Comment threads from news articles related to the topic

Two different tasks/topics
1. In the context of the war in Ukraine, are participants in the conversations more sympathetic towards Russia or Ukraine?
2. Are participants in the conversations more supportive of Hezbollah or ISIS?

User Study Scenario

For each comment
◦ Mark it as relevant to the topic or NTR
◦ If relevant, Use the pull-downs to "fill in the blanks" of a contextual note
◦ Provide a confidence rating
◦ Optional: any other important analyst comments

Not entirely realistic
◦ Analysts wouldn't work this granularly
◦ Necessary conceit to get enough data without overtaxing analysts

User Study Interface Mock-up

Pending pre-publication review

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas, Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 391

User Study Interface Mock-up

Pending pre-publication review

User Study Interface Mock-up

Pending pre-publication review